DOCUMENT RESUME

ED 227 672 FL 013 531

TITLE	Planning for Student Assessment: Leader's Guide. Bilingual Evaluation Technical Assistance, Workshop III.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
PUB DATE	[80]
NOTE	56p.; For related documents, see ED 221 565-569, FL 013 523-524, FL 013 526-529, FL 013 532, FL 013 534-535, and FL 013 537-538.
AVAILABLE FROM	Center for the Study of Evaluation, UCLA Graduate School of Education, Los Angeles, CA 90024.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Guides - Classroom Use - Guides (For Teachers) (052)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Bilingual Education Programs; English (Second Language); Evaluation Criteria; *Evaluation Methods; *Language Proficiency; *Language Tests; Limited English Speaking; Program Effectiveness; Program Evaluation; *Student Evaluation; *Workshops

ABSTRACT

This leader's manual for a workshop on planning bilingual student assessment describes the workshop, presents information necessary to provide instruction leading to workshop exercises, and provides step-by-step instructions for conducting the workshop. The workshop leader should become familiar with this manual as well as with the workshop text and the participant's handbook. The manual aids the leader in providing workshop participants with practice in planning bilingual education evaluations and with an overview of the issues that should be addressed during such evaluations. The topics include: (1) how to organize the workshop; (2) instructions for using the leader's manual; (3) identifying legal assessment obligations (participant list of program assessment needs based on legislation and funding source requirements, assessment in bilingual programs, and coordinating assessment needs); (4) what to test (categorizing tests and test selection); (5) initial screening of tests (basic terminology, two simulated test descriptions, and test selection); (6) using tests; and (7) final screening - comparing the technical merits of tests (review of technical terms, reading a printout, reading a technical manual, and text selection simulation). (NCR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227672

Bilingual Evaluation Technical Assistance

Workshop III

Planning for Student Assessment

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Leader's Guide

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CSE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL013531

Center for the Study of Evaluation

The mission of the Center for the Study of Evaluation is to conduct inquiry, from a variety of perspectives, into the nature of educational programs and services. Our commitment to inquiry into the field of evaluation grows from the belief that school practices and the competencies and satisfactions of those who participate in the educational enterprise can benefit from information collected in accordance with social science methodologies. Activities of CSE involve study of the instruments and methodologies for collecting information as well as the sociopolitical contexts of educational decisions as a means of contributing to the long-range growth in effectiveness of public education.

Information about CSE and its publications may be obtained by writing to:

Director, Public Information
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

Table of Contents

	Page
How to Organize the Workshop	1
Pre-Workshop Preparation Checklist	3
Room Arrangement	4
Instructions for Using the Leader's Manual	5
General Workshop Schedule	7
Introduction	8
Stage 1: Identifying legal assessment obligations	14
Activity 1: Participant list of program assessment needs based on legislation and funding source requirements	15
Activity 2: Assessment in Bilingual Programs	18
Activity 3: Coordinating Assessment Needs	20
Stage 2: What to Test	23
Activity 1: Categorizing Tests	24
Activity 2: Simulation: Test Selection	25
Activity 3: Follow Up On Simulation	25
Stage 3: Initial Screening of Tests	29
Activity 1: Basic Terminology	31
Activity 2: Initial Screening of Tests from Two Simulated Test Descriptions	32
Activity 3: Test Selection	34
Stage 4: Using Tests	35
Stage 5: Final Screening: Comparing the Technical Merits of Tests	36
Activity 1: Review of Technical Terms	37
Activity 2: Reading a Printout	37
Activity 3: Reading a Technical Manual	48
Activity 4: Decision Making in Context: Text Selection Simulation	48
Activity 5: What About My Bilingual Program?	49

HOW TO ORGANIZE THE WORKSHOP

Recruitment

The BETA workshops were designed for audiences with three levels of involvement with bilingual evaluation. At the most general level, all people involved in bilingual program, both bilingual and monolingual program teachers, students, parents, and community members will find some parts of the materials valuable. A second audience is educators directly involved in bilingual program planning and implementation who are familiar with many of the general principles of curriculum development and testing, but who have no special training in evaluation. The third group for whom the materials were designed is research and evaluation specialists who may not be well-acquainted with the special demands of bilingual program evaluation. Since the concerns of these three audiences are often very disparate, we suggest that you organize workshops to serve each group separately. Thus, workshops will serve one of the following categories of levels of expertise:

G: Workshops for all levels of expertise including parents and community.

S: Teachers, administrators and school site staff with instructional and administrative experience in bilingual programs, but no specialized evaluation training.

R: Evaluators and research staff familiar with design, measurement and implementation procedures used in program evaluation, but unfamiliar with the special demands of bilingual program evaluation.

Notifying Participants

Remind each workshop participant of the workshop about one week prior to the session. Emphasize that the workshop, unlike many conferences and meetings, will require their active participation. Encourage them to bring with them to the sessions questions and materials from their own programs that are related to bilingual evaluation. Explain that the workshops are designed as working and troubleshooting sessions for their own concerns and the intent is to provide immediate help in planning bilingual evaluations.

Planning Lunch and Refreshments

To maintain the workshop schedule, it is advisable that the participants receive lunch and refreshments in the workshop room, or in an adjacent room. A pre-arranged lunch menu will avoid the confusion and inevitable delays associated with placing individual lunch orders. If the sessions are to run during normal working hours, fifteen minutes can be spared for coffee breaks. Better yet, have coffee and refreshments available to the participants throughout the session. The latter approach seems to work better, if the participants are told politely that they should not spend excessive time around the coffee pot during the session.

Pre-Workshop Preparation Checklist

Use the checklist below to ensure that the proper materials are obtained for conducting the workshop:

The following items are the responsibility of the Workshop Leader:

_____Seating and workspace for workshop participants

_____A table for the Leader's materials and Overhead Projector

_____A blackboard for summarizing discussion activities

_____A room large enough to hold the required worktables and chairs for participants.

_____A workshop location that has lunch and snack facilities so that participants can eat together and take breaks with a minimum of time spent in travel.

The following items are supplied by the Workshop Distributor:

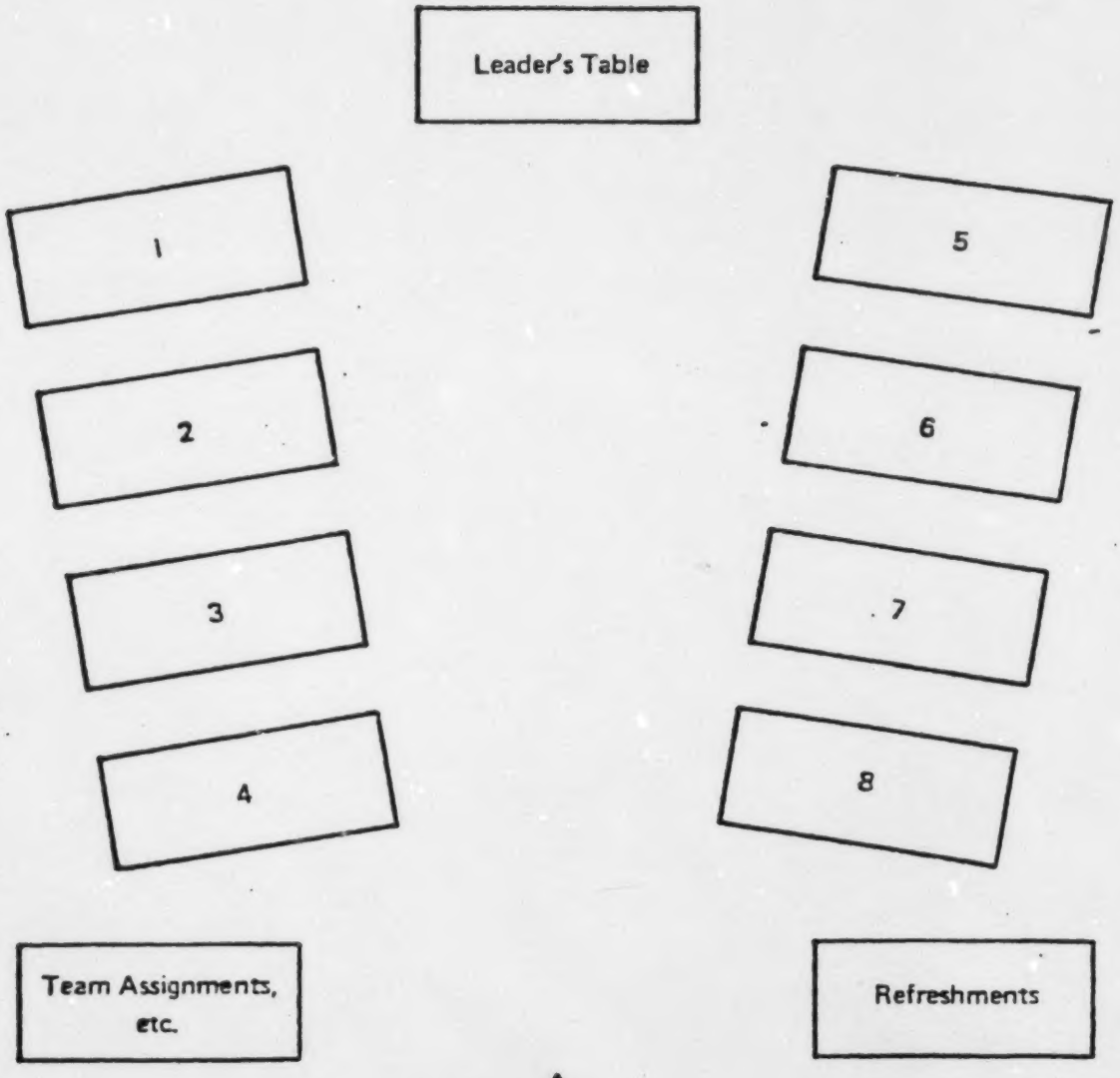
_____Leader's manual

_____One Workshop Text for each participant

_____One Participant's Handbook for each participant

Room Arrangement

A sample room arrangement for 24 participants (eight teams) appears below.



INSTRUCTIONS FOR USING THE LEADER'S MANUAL

This Manual describes the workshop, presents information necessary to provide instruction leading to workshop exercises and provides step-by-step instructions for conducting the workshop.

The Leader should become familiar with all workshop materials, Guide, Text, and Participant's Handbook, before attempting to conduct the workshop.

The major goal of the Bilingual Evaluation Technical Assistance materials is to provide practitioners with practice in planning bilingual evaluations and with an overview of the issues that must be addressed during such evaluations.

THE WORKSHOP LEADER SHOULD READ THE ENTIRE MANUAL BEFORE CONDUCTING THE WORKSHOP.

The Script, Extracts, and Directions to Leader

Each page of the Leader's Manual is divided into two parts. The left-hand portion of the page contains a blocked-off instructional "script." The following is an example of this script, which is read by the Workshop Leader:

Script

The purpose of this workshop is to introduce you to the basic concepts in bilingual evaluation.

If the Leader is unsure of the content of a workshop or with certain segments of it, she/he should adhere to the script. When scripted material is familiar to the Leader, she/he may wish to refer only to the extracts for guidance. These extracts appear on the right hand side of the script.

Script	Extract
...a design is a strategy which dictates from whom evaluation information will be collected, with what measuring device, and when data collection will occur..	Design Defined

Directions to the Leader consist of materials which appear between blocks of script. These directions may refer to such things as points the Leader should discuss, administrative materials, or how to organize an activity.

Script	Extract
You will have five minutes to read the materials in this section. Do not go on to the next section until I ask you to.	Background Information

(Directions) When the participants have read the section, answer any general or procedural questions based on the materials. After any discussion, go on to say:

GENERAL WORKSHOP SCHEDULE

This workshop has been designed to accomplish its objectives within four hours of instruction. Below you will find a general time schedule and the times allotted for each activity. Please detach this page for use during instruction.

	Time	Activity
Morning Workshop		
	8:00-8:20 AM	Coffee, introductions and general announcements
	8:20-8:40	Testing in Bilingual Programs: introduction
	8:40-8:55	Presenting Workshop Simulations and forming work units
	8:55-9:55	Identifying testing needs
	9:55	Break
	10:25-11:50	What subjects, skills to test?
	11:50	Lunch
Afternoon Workshop		
	12:50-1:45	What kind of instrument?
	1:45-2:45	Initial screening of tests
	2:45-4:00	Screening for technical quality
	4:00-4:45	Making recommendations for Assessment in your own bilingual program
	4:45-5:00	Participants' evaluation of Workshop III

Leader's Manual

Introduction

Testing in Bilingual Programs

An important aspect of a program's evaluation is the measurement and presentation of the students' level of performance. Student outcomes may determine whether changes are necessary in any of its components, such as its goals and objectives, instructional curriculum, or staffing. However, the accuracy and context within which student outcomes are presented depend greatly on the quality of the evaluation including the instruments and methods used for its implementation.

It should be noted that "assessment of student outcomes" is not synonymous with "testing". Much information can be obtained through other methods such as observations, use of archival data case studies. (For more information, see Workshop IV: Planning for Implementation Evaluation). This workshop focuses on testing for the assessment of student performance because testing issues are important in discussing the evaluation of bilingual program effectiveness. You will find that the workshop does not attempt to prescribe specific tests or methods for planning for student assessment. Because of the wide variety of program needs and characteristics, it would be improper to prescribe one method of planning for student assessment. This workshop will highlight basic questions which you should consider in the process of planning for student assessment:

1) For what purpose(s) will we test the students?

2) What are the characteristics of the students we need to test? What are their language levels? Should they be tested in English or their native language? What areas can they be tested in?

3) What skills or content areas need to be tested for each testing purpose?

4) What assessment instruments can/should we use? Are the tests we are using or we have used in the past acceptable? If not, how do we select tests? How do we know they will accurately and fairly assess what we need?

5) Once the test results have been obtained, what is the correct interpretation of the results? Should we use normal curve equivalents or grade equivalents? Do low scores on a standardized norm referenced test mean that the program isn't working? If not, what other factors should be examined?

LANGUAGE CLASSIFICATION AND ASSESSMENT

The most distinguishing factor of bilingual programs is that students enter the program with a range of skill levels both in their native language and in English. This results in various subgroups of students (as defined by their skill level in the first and second languages within each grade level as well as within each classroom). This poses a grave problem in the assessment of students in order to determine program effectiveness. Why? Because each of these 'subgroups' will progress at varying rates depending on the quality and quantity of instruction in the first and second languages as well as on the age and initial skill level of the students comprising the subgroups. Therefore, if the progress of all of the bilingual students in the program is measured as a whole, then the differences between subgroups will be blurred. Consequently, a program may not seem to be doing well as a whole; yet, when subgroups are considered separately, it could be that some groups are progressing at an acceptable rate and others are not.

If the evaluator does not use designs, instruments and assessment methods which accurately reflect what has occurred in the program, particularly with respect to the instructional curricula, evaluation may not accurately reflect student growth or student skill levels. This is particularly important in bilingual programs because student achievement is linked to their growth in first and/or second language acquisition.

WORKSHOP PURPOSE

The aim of this workshop is not to offer "answers" to all of the complex questions associated with testing in bilingual programs. Instead, it intends to point out key decision areas you are likely to encounter when planning the testing component of your evaluation and to offer strategies for making decision about testing.

The workshop hopes to instill awareness of the basic stages in the testing processes of an evaluation and of the roles of school personnel within this process. Within each stage you will participate in some hands-on learning experiences that will help you understand the assessment process in bilingual education. The workshop activities will provide practice in the following:

1. identifying the assessment needs of your school or program.
2. identifying and describing the types of measures which are used by your district, school or program to meet these assessment needs.
3. comparing and examining some of these instruments for technical adequacy, curricular relevance and appropriateness for the student populations in your school(s).
4. making recommendations regarding the assessment process in your school or program.

You might want to refer again to Figure 1 to see how these assessment planning activities are interrelated.

5. record keeping

In order to achieve these objectives, we will actually go through each stage and perform some of the main activities required by each part of the process and discuss problems. We will manipulate real program information and data. You have the option of using information from your own schools and/or programs, or of using our simulated program descriptions.

Using this information we will go through the stages of the testing process. Please refer to Figure 1. The personnel listed above each box are examples of how some school districts and programs allocate their personnel. IT IS NOT MEANT TO BE PRESCRIPTIVE. As a matter of fact, as we discuss each "box" you should modify this to fit your program situation.

WORKSHOP ORGANIZATION

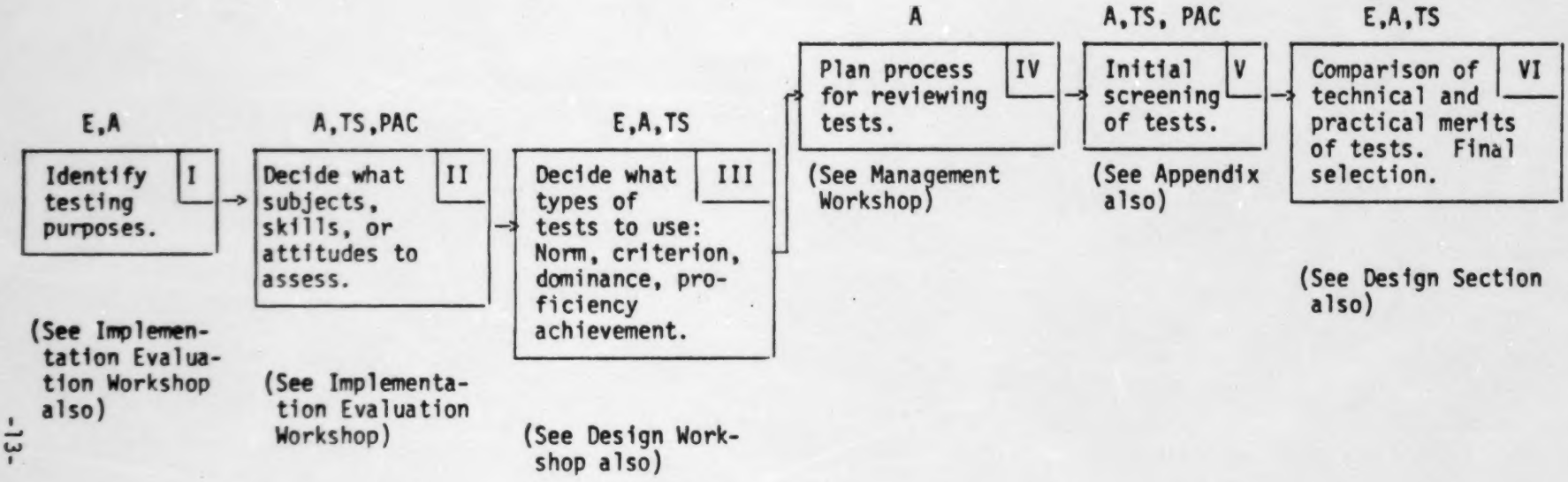
(Leader)

At this point, read the simulation (or have audience read it). This is so that the participants will know what kinds of information they should have on hand.

OR

Find out how many people want to use the simulated descriptions and give hand-outs.

Figure 1


E,A
Identify testing purposes. I
(See Implementation Evaluation Workshop also)
A,TS,PAC
Decide what subjects, skills, or attitudes to assess. II
(See Implementation Evaluation Workshop)
E,A,TS
Decide what types of tests to use: Norm, criterion, dominance, proficiency achievement. III
(See Design Workshop also)
A
Plan process for reviewing tests. IV
(See Management Workshop)
A,TS, PAC
Initial screening of tests. V
(See Appendix also)
E,A,TS
Comparison of technical and practical merits of tests. Final selection. VI
(See Design Section also)

E = evaluator
A = administrator
TS = teaching staff
PAC = parent advisory committee

Stage 1

Identifying Testing Needs

In the first section of this workshop we will examine the five major functions of student assessment more closely. You will examine what each assessment function stipulates in terms of: 1) content areas and skills; 2) types of tests used; 3) test scheduling and, 4) whether group or individual student assessment questions need to be answered. We will then identify the overlaps among the testing functions in order to minimize the amount of testing by maximizing the utility of each testing effort. In the final exercise of this section you will map out a testing plan.

Let's first look at assessment obligations which are shaped by federal and state legislation, as well as by the contractual stipulations of your funding sources.

ASSESSMENT NEEDS AND OBLIGATIONS

Activity #1: Participant List of Program Assessment Needs Based on Legislation and Funding Source Requirements

(Leader)

1. Find out what types of federal funding participants have to deal with. For example: TVII, TI, TIV or combinations (See appendix for a summary of assessment requirements for each funding title).
2. Participants should indicate what their state funding testing requirements are. Or, if they do not know, refer to Rizizarry's Review of the Legislation. Those using simulated descriptions should read the paragraph describing its federal and state requirements.
3. Allow 15 minutes for participants to read materials describing these testing obligations.

4. Write out on blackboard or overhead, a simplified version of the requirements as stated by the participants.

State the requirements in as simple a fashion as possible. For example TVII funds require:

- Description/documentation of placement of children in the program.
- Comparing the performance of participating children on tests of reading skills in English and in the native language to be used in the program with an estimate of what such children's performance would have been in the absence of such a program.
- Pre-and post-test results on reading tests using mean scores and appropriate tests of statistical significance.
- A description of measurement instruments to be used by the program, including a rationale for selecting these instruments and procedures to be followed in their use.

Note that almost all of the above requirements seek information which is also useful for program improvement, and operations. For example, language proficiency and reading.
Also keep in mind that the requirement of "comparative data" does not necessarily mean that norm-referenced tests must be used. "Comparative data" can be obtained through other means.

(Leader)

5. After participants have listed the federal and state requirements, look at Figure 2 using the overhead. Have them determine where these requirements overlap. For your convenience, the symbol Ⓓ and underlining have been used to indicate apparent overlaps.

Figure 2

Sample: Participants' List of Federal & State Requirements

Federal		*State
1. Documentation of student placement in the program		1. Not mentioned
2. Normative scores on achievement tests of reading and math in English and, where possible, on native language	ⓓ	2. Achievement scores in reading and math
3. Test scores showing progress of student attainment of English	ⓓ	3. Scores showing student progress in English reading and speaking skills
4. Specified context: English reading, math, other content areas in English or native language, wherever possible.	ⓓ	4. Reading and listening comprehension, writing and speaking and math in English. Reading and other content areas in the native language
5. Overall student progress	ⓓ	5. Overall student progress

*Student assessment practices vary from state to state. However, most states do require some documentation of the placement process as well as of students' progress in achievement.

Activity 2: Assessment in Bilingual Programs

In order to minimize over-testing and identify both tests and testing efforts which can be used for multiple purposes, you might map the program testing requirements. One example of a map to coordinate testing requirements appears in Figure 3. By planning testing for several purposes, you can guard against over-testing. For example, a comprehensive diagnostic tool used for language proficiency assessment can also provide information for: 1) program placement, 2) diagnosis for instruction, 3) progress monitoring. Or, you could use state-mandated achievement test results to report required test scores to funding agencies. You might also consider reviewing student achievement information from other classes in which they are enrolled as part of progress monitoring, exit testing, or evaluation.

TEST

Figure 3

Types of Assessment	Instructional Programatic Decisions	Factors being Assessed	Language of Measurement
Language Assessment	1. Entry 2. Instructional Placement 3. Exit	Listening Comprehension Speaking Fluency Reading Comprehension Writing	L1 & L2
Achievement	1. Monitoring students learning 2. Program Evaluation	Language Arts Math Social Studies Sciences	Language in which instruction is given
Special Bilingual Education Programs	1. Psychological Diagnoses 2. Instructional Placement	IQ Tests	Students strongest language

(Leader)

A. Have participants look at both lists of assessment needs (see Figure 3) and underline (or separately list) those skills, content areas or other aspects which are specifically mentioned. (For example: Reading comprehension in English/native language; math skills in English/native language attitudes toward instruction given in English, etc.)

B. Have participants look at their curricular objectives to identify any other area which needs to be assessed. Participants using the simulation materials should look on pages of their workbook.

Pay close attention to the relationship between the purpose for assessment, and the type of information needed about student language skills. Try to answer the following:

- What does a language dominance test measure?
- What does a language proficiency test measure?
- What is achievement testing?
- Is there any difference between proficiency tests and achievement tests as measures in language skills?

Refer to language skills Figure. Think about which of these skills/attitudes are important to assess for each of the five assessment purposes: placement; instruction; exit; achievement and evaluation and discuss.

(Leader)

Refer participants to Section I-II of the text

Activity #3: Coordinating Assessment Needs

(Leader)

Have participants prioritize assessment needs and formulate a single list for their programs. Allow 15 minutes for this and then open a discussion session.

(Leader)

1. Using Figure 4, instruct participants to write when each assessment should occur.
2. Once Figure 4 has been filled, participants should study the figure to determine where the overlaps are in terms of skills tested, type of test used and time tested.
3. Feedback:
   1. Work with the participants to complete and circulate the sample plan.
   2. When this is done, show on the overhead. Figure 4, Mapping Out the Testing Plan, is an example.
4. Work with participants to make recommendations to reduce the amount of testing.

Figure 4

Mapping Out The Testing Plan

PURPOSE	SKILLS TESTED	TEST SCHEDULE
PROCESS: Initial eligibility of students.	Language proficiency or dominance. Language skills in native language and English.	Beginning of each semester, or at least every fall semester.
PROCESS: Further testing of intermediate & advanced LEP for placement within the bilingual and/or monolingual programs.	Further tests in English language skills (either proficiency or achievement). Tests in other subject areas (i.e., math) either in English or native language.	Beginning of the semester.
PROCESS: Classroom teacher assessment of students for placement in appropriate reading, math or other subject group.	May be an informal appraisal by teacher. Previous test scores may be used or a teacher developed test.	Beginning of school year or of each semester if student(s) is new to the classroom or program.
PROCESS: Exit/transfer of students either within the program or out of the program.	Assessment of English language skills using either a proficiency or achievement test. Possible assessment of skills in other subject areas in English.	End of each semester.
EVALUATION: Progress monitoring of students.	Usually daily or continual appraisal by teacher through classroom tests, etc. Some schools do additional testing in subject areas such as science & social studies.	Continuing (by teacher) or annually (if by school district).
EVALUATION: Achievement testing for funding of other purposes.	Standardized achievement tests in reading and math in English and/or the native language.	Either pre/post at beginning of fall & end of spring semesters or, annuually in spring semester.

Stage 2

What to Test

Activity #1: Categorizing Tests

(Leader)

1. Have participants quickly read pages 6-8 and 19-28 of the text.

2. Aid participants in categorizing assessment instruments they use in their programs according to whether they are: a) formal or informal instruments; b) norm referenced, criterion referenced; c) language dominance, language proficiency; or d) achievement tests.

3. Have participants list first names on worksheet.

4. Have participants check or fill in appropriate test descriptions on their worksheet.

5. As you go through each category, ask participants to describe the categories on the chart. The following questions could be used to lead the discussion:

    1. What is the difference between formal and informal assessment?
    2. What standards to norm referenced and criterion referenced tests use to measure student performance?
    3. What is "norming"?

If participants are having difficulty with concepts presented in the category chart, refer them to pages 19-28 in the text for a detailed discussion of these principles.

Allow the participants time to review the basic concepts of testing.

Follow up the reading with the following discussion.

4. How can you tell whether a norm referenced test has been normed with a population which is comparable or similar to your student population?

5. Can you "norm" a criterion referenced test?

6. How can you tell whether a CRT is appropriate for your student population?

Activity #2: Simulation: Test Selection

(Leader)

1. Have participants turn to vignettes and answer sheets for page 2 - Activity 2 in their handbooks.

2. Divide audience into groups of 4-6 and assign each group responsibility for developing testing strategy for the vignette.

3. Each group must:

   a) identify at least three evaluation needs, one for summative and two for formative evaluation.

   b) determine which needs could be fulfilled by testing

   c) determine what kind of information (CRT/NRT/survey/diagnostic) would be <u>ideal</u> in each case and which would be <u>feasible</u> given:

      - the availability of instruments
      - size of the population
      - population characteristics
      - deadline for the evaluation

4. They should fill in the checklist - with this information.

Activity #3: Follow Up On Simulation

(Leader)

1. Ask participants to share their testing strategies and some of the problems they came across.

Overhead:

Categories of Assessment Instruments

Name of Test	Test Use		Score Interpretation		Content		
	Standardized	Other	NRT	CRT	LP	LD	Other
CTB Espanol	x		x				rdg. comp. achievmt.
SRA End of unit reading test	x			x			decoding/ encoding
Student class participation		x		x			pronuncia-tion
Weekly spelling test		x		x			reading/ vocabulary

Simulation: Follow-up

Questions

1. What conflicts did you find between ideal and feasible testing needs?
2. How do assessment requirements for the different funding sources overlap in terms of:
    a. content to be tested (i.e., reading in English, math)
    b. testing design (i.e., pre/post or annually)
    c. dates of testing (i.e., fall & spring or once a year)
    d. number/groups of students to be tested (i.e., all the students in the school or bilingual program students only).
3. How would you handle achievement testing for the LEP/NEP when you must serve several language groups? Recall that the program services 3 language groups, plus ESL for Vietnamese. (Below are some of the options if the participants have difficulty).
    a. Test all students with the same achievement test in English, (except NEP who are not sufficiently proficient in English). and give each language group a parallel or similar test in their native language.
    b. Test only those students who are sufficiently proficient in English on the English achievement test. Give others an achievement test in their native language or no achievement test at all.
    c. Test only those students in the bilingual program who are near the transitional stage, in English.

d. Test advanced LEP students in English only and other students in their native language only.

4. How would you obtain normative data in subject areas subject areas such as science or social studies?
5. How would you measure the affective objectives of self concept and cultural pride?
6. What are some of the standards of comparison you could use in order to determine whether the student could "survive" or perform at the average or better level outside of the bilingual program?
7. What are some ways of following up transitioned students to see whether they are performing satisfactorily in the monolingual English program?
8. What could/would you do if some of the students were not performing satisfactorily after they have been mainstreamed?

Stage 3

Initial Screening of Tests

INFORMATION SOURCES OF TESTS

Once you have determined what you need to test and the types of data that are needed, the next step is to gather a list or description of actual tests which potentially fulfill these general characteristics. Where would you find such tests? There are several key resources of information. Page 17 in your text lists catalogues of available instruments.

Publishers of educational materials would also be a good source. There are several federally funded research and development centers, dissemination centers and training centers which may have information regarding bilingual materials. Furthermore, the federal government has a Bilingual Clearinghouse Hotline (800) 336-4560. A call to your state or city office of bilingual education may also be helpful. Local universities and colleges may provide other sources of information or have tests on hand. These sources also provide technical assistance if you intend to develop your own tests.

What kinds of features should be considered for a preliminary screening of tests? Let's quickly go over some of these features. Pages 29-41 of the text discuss these test features in more detail. We also have an exercise coming up which will illustrate their importance.

Activity #1: Basic Terminology

(Leader)

Depending on the sophistication and/or interest of your group, spend time describing the salient aspects of each test feature. Encourage audience to share their experiences in this aspect. Refer to the Test Selection Terminology in the Participant's Handbook.

Features to be discussed:

- validity (general)
- linguistic and cultural appropriateness
- format and test directions
- administrator qualifications & training
- test leveling
- time
- cost

1. Write each item on the board
2. Ask participants for definition
3. If participants have difficulty, allow them time to read the Test Selection Terminology; then begin an activity.

Activity #2: Initial Screening of Tests From Two Simulated Test Descriptions

(Leader)

1. Have participants turn to mock-ups of two language assessment tests in their Handbooks.
2. Have participants break into small groups and prepare answers for the questions below.

* Is this a language dominance, proficiency or achievement test?

(Test A is a dominance and "lightweight" proficiency test good for initial placement. Test B is a more in-depth proficiency test, good for diagnostic and placement purposes.)

* What language skills do the tests measure and how?

(Test A measures only oral and aural skills. It also claims to measure vocabulary and grammar within this context. Students are shown a picture and asked questions regarding the pictures. Some of the questions are "elicited response" items, that is, they require a particular language structure in the response. Other questions are "open ended" and allow the student to respond with some creativity and spontaneity.

Test B measures reading, writing, listening and speaking skills. For the listening and speaking skills, students are asked specific questions, some of which require a short or long answer. The reading and writing tests are group administered and include specific response items and dictation as well as a less structured writing assignment.)

* How easy is it to administer these tests? How long do they take?

(Test A is relatively easy to administer and requires a bilingual test administrator who is familiar with the students' language and culture, particularly their regionalisms. It takes about 20 minutes to administer both the English and native language version of the test for each student.

Test B administration requires some training because the test is long and involves handling of a tape recorder or writing down student responses for the listening and speaking subtests. The writing and reading group subtests are easy to administrate. Total testing time is 40 minutes for the oral and listening skills subtests in both languages per student. It takes one hour per group to administer the reading and writing subtests.)

* How easy is it to score these tests?

(Test A requires two kinds of scoring: 1) On-the-spot scoring of each response by the test administrator; 2) Total scoring and categorization of student at a later time. The responses are weighted differently according to the difficulty of the question. There are some problems with the on-the-spot ratings in that the raters must be extremely proficient in the student's native language and about the regionalisms of the student's language to insure an accurate and fair rating. Furthermore, each response must be written and/or recorded. Sometimes student responses will be much longer than those required by the test and the rater will have to use his/her judgment as to the acceptability of each response.)

There is no mechanism for checking the reliability of raters. The test scores are fine for reflecting general language dominance and proficiency. However, the distinction between the more advanced bilinguals is not really clear.

Test B: All of the four categories (subtests) of this instrument are weighted equally; however, each subtest has weighted responses. Therefore, the evaluator or other school personnel should decide which subtests (if any) are based on all four subtests, or only some. Because of this system, the total scoring can be complex. Raters do not do on-the-spot scoring of student responses, since the oral part is recorded. Two or more raters must listen to each student's tape and score responses. Raters' scores on each student must agree. This procedure insures inter-rater reliability.)

* How cost-effective are these tests?

(Test A is not very expensive ($16.00 per packet of 35

plus an examiner's manual) and takes relatively little time to administer for an individualized test. However, the information it renders is best suited for initial placement based on speaking and listening comprehension skills only.

Test B is more expensive, $45.00 per packet of 35, for all four subtests. However, this does not include the cost of the tape cartridges, nor does it include tape recording equipment. This test does take longer to administer. However, it obtains information which could be used for placement, diagnostic, monitoring and transference purposes as well.

The cost-effectiveness of these instruments depends on your program's assessment needs (both short and long term) and on your deadlines. Only you and your colleagues can make this decision.

Activity #3: Test Selection

(Leader)

Thinking of the vignettes you just worked with, which of the two instruments would you have selected to include as part of the assessment criteria for placement and/or transference?

Stage 4

Using Tests

Stage 5: Final Screening: Comparing the technical merits of tests

(Leader)

The tone and activities of this section should be especially geared to the interest and level of expertise of the participants. If all or most of the terminology is unfamiliar to the participants, then go over the explanations on the pages of the text.

Recall that you categorized the tests used by your district (or simulated program) according to whether they were informal or informal instruments, norm or criterion referenced. You also considered the general testing purposes they might serve. Since you have had the opportunity to examine some hypothetical language assessment tests, we will now scrutinize some hypothetical achievement tests. We will pay particular attention to how their results can be presented and interpreted. The results will be given to you in mock ups of the printout which are usually generated by test companies in reporting test results; you will also have a mock-up of a technical manual.

We will discuss the meaning and application of each item on the printout. As we go through this, please take "mental notes," think of your program needs (whether it's the simulated or the actual program) for assessment. At the end of this discussion, you will do the following:

1. Examine the instruments used by your district for probable reliability and validity, and other aspects of technical merit;
2. Determine whether these instruments appropriately fulfill either your district or simulated program needs.

OR

3. Using the three vignettes handed out previously, you will identify the types of assessment information which would best suit the needs of each program.

Activity #1: Review of Technical Terms

(Leader)

1. Pass out printout and technical manual mock-ups and set up your printout mock-up on the overhead.
2. Ask participants the meaning of the items which appear on the printout as you go through it.

- Concepts for discussion:

  types of information produced by norm referenced tests;
  types of information produced by criterion referenced tests;
  concepts listed in the workshop outline

Activity 2: Reading a Printout

The content of the printouts is usually determined by what the school district has requested. They will generally contain raw and transformed scores for each student. It may also give percentile rankings, grade equivalents or normal curve equivalent scores for each classroom, grade level in the school or program. For school

districts, totals may be obtained for each school, or breakdowns according to language proficiency for the district and/or each school. We will see examples of these breakdowns after we have gone through the other items on the printout. Remember that this printout is only a simulation of those which are commonly put out by testing companies. There are, of course, different formats and varying types of information which appear on the printout, depending on the services provided by the company and the district needs. Because there is a variety of formats and information provided by these printouts, the first thing you have to do is to acquaint yourself with where the basic information is provided, such as:

READING A PRINTOUT

1. Whether the printout provides information by classroom, school, or district. (Leader: point this out on the printout.)
2. What subtest scores are provided, and whether each subtest has a total score, or whether there is a total score for the complete battery. (Leader: point this out on the printout.)
3. Which test level and form was used? Did the district do out of level testing? If so, then the score interpretations may be different. (Leader: indicate this on the printout.)

4. The next source of information you need is the key to the printout scores. This usually appears at the bottom or top of the actual scores in a format similar to the one presented here. Let's go through each item on this key.

(Leader)

1. Ask participants to explain the meaning and use of each item in the key. In the case of printout #1, the items are:

   "n count"
   raw score, national percentile, scale scores, grade equivalent scores.
2. If the participants do not know the meaning and use of each of the above items, provide them with the glossary definition and review it with them.
3. After reviewing the classroom printout, put the printout #2, "Evaluator's Listing by District." Review the following items:

   Mean, median, standard deviation. Explain the interpretation of the items in Section 4 of the printout.
4. Present the School District summary by language proficiency breakdown:

We will now look at another type of summary which can be very useful for evaluation purposes. This is a summary which shows student performance according to different language proficiency categories. This is a particularly important data breakdown to have because, as you all may know, the score

READING A PRINTOUT

reports for the district or school as a whole may be misleading since the language proficiency of the students who took the test is not reflected. Therefore, a whole school program or bilingual program may appear to be unsuccessful because the lower scores of NEP/LEP, who may not be ready to take an achievement test in English, will bring down the total scores. By having a breakdown such as this, the evaluator, administrator, or teacher can easily see how each group of students is doing within the school or program, within the district, and as compared to either national or local norms. This summary is provided by some of the testing companies if the district requests it. If the testing company which provides your tests does not have this service, you can ask your evaluator to do a school and district breakdown.

(Leader)

1. Give participants a few minutes to look at the Summary. Then open a question/answer period about the three printouts you have just reviewed.

In order to understand the true significance of these figures, you will have to look at the technical manual. The technical manual will explain how the reliability and validity of the test was determined, as well as other test development procedures such as: sampling; item selection; norming; content selection. Scores produced by the sampling norm

populations are often included. If the test used word frequency counts, objectives or other information found in reading programs, the manual may indicate which reading programs were used. The manual will also explain the procedures followed to obtain transformed scores. We will now look at the simulated technical manual.

(Leader)

1. Pass out the manuals and put the Table of Contents in the overhead. Some participants may have brought their own manuals. In this case, they should look up those items which are being discussed in them.

Please turn to the Table of Contents. Notice that some of the items have an asterisk. The manual will contain only those items. The reason is that the remaining items, such as the rationale and content selection, are either self-explanatory or have little or no bearing on this discussion. There are basic features of test development for which the manual should provide information.

1. How the test items were developed and or selected.
2. What types of populations were used for standardization.
3. What the standardization procedures were.
4. What types of scores and reports are provided.
5. How the test should be administered and scored.

6. What kinds of services the test company provides, at what cost, and what the turnabout time is.

For our immediate purposes, we will only discuss the first four features.

(Leader)

1. Have the participants turn to the Technical Manual section on item development and selection. Have them read this section and ask the following questions:

1. Is there anything in this section which you do not understand?
   (There may be some questions about the meaning of point biserial correlation coefficients.)
2. What kind of information is provided about item development and selection? (Leader: The following are some possible answers. Participants may add others.)
   a. Who writes and/or selects the items and objectives?
   b. What are the criterion for test writing and/or selection?
   c. How the items and objectives were validated.
   d. Statistical criterion for item selection.

3. What other kinds of information would you need to know?

   (Leader: The following are some possible answers. Participants may add others.)

   a. The characteristics of the pilot test population.

   b. The specific reading programs and other curriculum material on which the items were based. Were any of the materials used in bilingual programs?

   c. Were any of the teachers experienced and/or knowledgeable about second language acquisition and bilingual education?

   d. How were items screened for bias?

(Leader)

After you have finished discussing the section on item selection and writing, discuss the following questions:

1. What are the characteristics of the populations used for norming in terms of:

   a. number of students

   b. number of grade levels

   c. geographical location of the norming populations

   d. ethnic characteristics of the norming populations

   e. language proficiency

2. How and why were these schools selected?

NORM GROUPS

3. How many times were they tested and when?

4. What was the test-retest reliability of the instrument at each level for

   a. 3 to 6 month test/retest

   b. 2-3 week test/retest.

5. Which one would you look at in selecting a test for program evaluation when you are using the same test for pre-and post-test observation?

   (Answer: The 6 month reliability coefficient.)

6. Which one would you select when reporting the reliability of scores on one occasion in calculation of the standard error of measure for groups of students?

7. Which would you choose for making decisions about individual student achievement?

   (Answer: None - use standard error types of reliability. These are scores which are rarely, if ever, reported.)

REVIEWING TECHNICAL MANUAL

(Leader)

After you have discussed the above questions with participants, open a short question/answer period.

PRINTOUT NO. 1

CLASS SCORE SHEET

TEACHER:

SCHOOL:

SCHOOL ADDRESS:

TEST DATE:

CLASS GRADE:

TEST: THE BILINGUAL LANGUAGE AND MATH SKILLS TEST

TEST LEVEL: 2

TEST FORM: A

KEY TO PRINTOUT ⟶ (RAW SCORE) RS | SS (SCALE SCORE)
(NATIONAL PERCENTILE) NP | GE (GRADE EQUIVALENT)

STUDENT NAME	BIRTHDATE	LISTENING COMPREHENSION		TOTAL	READING COMPREHENSION	TOTAL	MATHEMATICS		TOTAL	TOTAL BATTERY
		13 | 331 38 | 27	35 | 481 92 | 59	48 | 405 80 | 43	16 | 376 62 | 36		26 | 338 58 | 33	6 | 303 28 | 24	48 | 335 52 | 31	
		20 | 385 71 | 39	14 | 380 45 | 29	34 | 367 61 | 34	11 | 336 38 | 27		16 | 309 30 | 28		27 | 243 14 | 21	
		21 | 390 73 | 40	12 | 340 35 | 26	33 | 364 39 | 34	18 | 404 76 | 43		9 | 285 14 | 22	11 | 358 57 | 34	38 | 313 34 | 28	
		13 | 331 38 | 27	14 | 360 45 | 29	27 | 333 41 | 28	10 | 292 18 | 23		4 | 247 5 | 12	3 | 246 8 | 16	17 | 236 2 | 14	

PRINTOUT NO. 2

DISTRICT:

GRADE:

EVALUATOR'S SUMMARY LISTING BY DISTRICT

TEST LEVEL: 7

LINE		LISTENING COMPREHENSION	READING COMPREHENSION	READING TOTAL	MATH COMPUTATION	MATH APPLICATION	MATH TOTAL	BATTERY TOTAL
1	N COUNT	1328	1341	1271	1284	1221	1175	7
	MEAN GE	8.4	8.6	8.5	9.7	8.6	9.1	6.8
	MEDIAN GE	8.1	8.3	8.2	9.2	8.2	8.8	
	SGE	8.4	8.4	8.3	9.0	8.4	8.9	6.7
	STD DEV GE	2.7	2.9	2.7	2.5	2.6	2.4	3.1
	MEAN SS	541.5	552.5	545.4	579.0	546.1	560.3	493.6
2	MEDIAN SS	535.3	550.1	542.3	581.4	541.1	558.3	
	STD DEV SS	74.1	74.5	75.6	77.5	72.9	70.5	82.4
	MEAN NCE	42.9	44.3	43.2	48.7	42.7	45.4	29.7
	MEDIAN NCE	41.3	44.1	42.4	49.3	41.2	45.1	
	SNCE	43	44	43	49	43	45	30
	STD DEV NCE	18.5	18.4	18.8	18.6	18.3	17.8	21.3
	MEDIAN NP	34.3	38.2	35.4	48.4	34.1	40.3	
	SNP	37	39	37	47	36	41	17
	75%ILE (Q-3)							
	GE	10.3	10.7	10.5	12.5	10.2	11.1	
	SS	587.4	599.1	594.0	631.1	592.2	606.3	
	NCE	54.4	56.1	54.9	61.1	54.3	56.6	
	NP	58.4	61.1	59.9	70.1	58.2	63.3	
	50%ILE (Q-2)							
	GE	8.1	8.3	8.2	9.2	8.2	8.8	
3	SS	535.3	550.1	542.3	581.4	541.1	558.3	
	NCE	41.3	44.1	42.4	49.3	41.2	45.1	
	NP	34.3	38.2	35.4	48.4	34.1	40.3	
	25%ILE (Q-1)							
	GE	6.4	6.3	6.5	7.8	6.6	7.4	
	SS	490.7	502.0	495.9	523.6	494.2	509.6	
	NCE	30.1	32.0	30.9	35.1	30.0	32.6	
	NP	17.1	19.0	17.9	23.6	17.3	20.0	
	N / QUARTER							
	76-99	159	185	159	243	152	164	1
	51-75	285	300	261	392	249	293	0
	26-50	373	415	395	309	368	331	1
4	1-25	511	441	456	340	452	387	5
	% / QUARTER							
	76-99	12.0	13.8	12.5	18.9	12.4	14.0	14.3
	51-75	21.5	22.4	20.5	30.5	20.4	24.9	0.0
	26-50	28.1	30.9	31.1	24.1	30.1	28.2	14.3
	1-25	38.5	32.9	35.9	26.5	37.0	32.9	71.4

PRINTOUT NO. 3

SCHOOL: SCHOOL SUMMARY TEST:

DISTRICT: TEST LEVEL:

STATE: SUBTEST: MATHEMATICS GRADE:

NON ENGLISH PROFICIENT

MEDIAN AND QUARTILE VALUES

	SCHOOL		STATE		DISTRICT	
	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %
3RD QUARTILE	72.0	35				
MEDIAN	49.0	15				
1ST QUARTILE	48.0	14				
NUMBER	2					

DISTRIBUTION

NATIONAL PERCENTILES	SCHOOL NUMBER	SCHOOL %	STATE NUMBER	STATE %	DISTRICT NUMBER	DISTRICT %
AT OR BELOW 75	2	100.0				
AT OR BELOW 50	2	100.0				
AT OR BELOW 25	1	50.0				

COMBINED LIMITED AND FLUENT ENGLISH PROFICIENT

MEDIAN AND QUARTILE VALUES

	SCHOOL		STATE		DISTRICT	
	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %
3RD QUARTILE	76.4	39				
MEDIAN	59.4	23				
1ST QUARTILE	46.5	13				
NUMBER	137					

DISTRIBUTION

NATIONAL PERCENTILES	SCHOOL NUMBER	SCHOOL %	STATE NUMBER	STATE %	DISTRICT NUMBER	DISTRICT %
AT OR BELOW 75	133	97.1				
AT OR BELOW 50	124	90.5				
AT OR BELOW 25	70	51.1				

LIMITED ENGLISH PROFICIENT

MEDIAN AND QUARTILE VALUES

	SCHOOL		STATE		DISTRICT	
	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %
3RD QUARTILE	73.2	36				
MEDIAN	59.2	23				
1ST QUARTILE	40.2	9				
NUMBER	51					

DISTRIBUTION

NATIONAL PERCENTILES	SCHOOL NUMBER	SCHOOL %	STATE NUMBER	STATE %	DISTRICT NUMBER	DISTRICT %
AT OR BELOW 75	51	100.0				
AT OR BELOW 50	49	96.1				
AT OR BELOW 25	26	51.0				

FLUENT ENGLISH PROFICIENT

MEDIAN AND QUARTILE VALUES

	SCHOOL		STATE		DISTRICT	
	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %	RAW SCORE	NAT'L %
3RD QUARTILE	79.3	44				
MEDIAN	59.5	24				
1ST QUARTILE	46.7	13				
NUMBER	86					

DISTRIBUTION

NATIONAL PERCENTILES	SCHOOL NUMBER	SCHOOL %	STATE NUMBER	STATE %	DISTRICT NUMBER	DISTRICT %
AT OR BELOW 75	82	95.3				
AT OR BELOW 50	75	87.2				
AT OR BELOW 25	44	51.2				

Activity #3: Reading a Technical Manual.

(Leader)

Have participants examine the technical manuals of the tests used by their districts (or mock-up manuals) to determine whether they have sufficient information about the validity, reliability and other technical aspects of the tests.

Activity #4: Decision Making in Context: Text Selection Simulation

There are two alternatives for this activity:

A. Have participants decide whether these instruments appropriately fulfill either their district or simulated program needs.

1. Participants should briefly write a rationale or list the reasons for their decision. Allow 10-15 minutes for this.
2. At the end of 15 minutes, randomly select participants who wish to share their decisions with the rest of the group.

(- OR -)

B. Participants should reconvene in their prior small groups. Using the vignette program description they worked on previously, each group should identify the types of assessment information which would best suit the needs of the program in terms of:

1. Criterion and/or norm referenced achievement test. Specify the content areas.
2. Would you want raw scores, stanines, percentiles, grade equivalents or other information?
3. What type of "breakdown," i.e., classroom, by school, by language levels, etc. - would you want?
4. If there is more than one non-English language group involved or more than one type of program, would all students or programs receive the same tests?

Give the groups about 20 minutes to formulate short answers to these questions. An answer does not necessarily mean a solution to each problem. If more problems are presented by these questions, given the nature of their vignette, then a clarification of the added problems would be fine.

NOTE: Worksheets are provided in the participant's workbook for this.

Activity #5: What About my Bilingual Program?

Ask participants to think of their own schools, programs or districts and how the whole process of student assessment is carried out as they see it. Ask them to formulate recommendations about each of the areas listed below:

1. Determining what specific skills, content areas to assess and when to assess these.
2. Determining the types of comparisons to be made.
3. Determining the tests to be used.

4. Test administration

5. Interpretation and reporting of test results.

Have the participants share their recommendations on the problems involved in implementing these recommendations.

Final Activity: Evaluation of this workshop by participants.

AFTER X MINUTES OF DISCUSSION SAY:

Before we close, I will pass out participant evaluation form of this workshop. Please help us to improve the workshop by taking a few minutes to give us your evaluation of the workshop.	PARTICIPANT EVALUATION